WHAT IS CLAIMED IS:

1. A computer-implemented method for processing a plurality of toponyms, said method comprising:

    in a large corpus, identifying geo-textual correlations among readings of the toponyms within the plurality of toponyms; and

    for each toponym selected from the plurality of toponyms, using the identified geo-textual correlations to generate a value for a confidence that the selected toponym refers to a corresponding geographic location.

2. The computer-implemented method of claim 1 further comprising using the confidences generated for the plurality of toponyms to rank documents according to their relevance to a search query.

3. The computer-implemented method of claim 1 further comprising selecting a set of initial values for the confidences for the plurality of toponyms, and wherein using the identified geo-textual correlations to generate values for confidences involves modifying the set of initial values based on the identified geo-textual correlations within the corpus.

4. The computer-implemented method of claim 3 wherein selecting the set of initial values for the confidences for the plurality of toponyms involves using a method of uniform priors.

5. The computer-implemented method of claim 1 wherein identifying geo-textual correlations involves identifying within documents in the corpus toponyms that have associated geographic locations that are nearby to each other.

6. The computer-implemented method of claim 1 wherein identifying geo-textual correlations involves identifying spatial correlation among geographic references of toponyms that are in textual proximity.

7. The computer-implemented method of claim 6 wherein textual proximity means within the same document.

8. The computer-implemented method of claim 6 wherein textual proximity means within the same document or any document closely linked with said same document.

9. The computer-implemented method of claim 1 further comprising processing the corpus by a named entity tagger prior to identifying the geo-textual correlations.

10. A computer-implemented method of generating information useful for ranking a document that includes a plurality of toponyms for which there is a corresponding plurality of (toponym,place) pairs, there being associated with each (toponym,place) pair of said plurality of (toponym,place) pairs a corresponding value for a confidence that the toponym of that (toponym,place) pair refers to the place of that (toponym,place) pair, said method comprising:

for a selected (toponym,place) pair of the plurality of (toponym,place) pairs,

(1) determining if another toponym is present within the document that has an associated place that is geographically related to the place of the selected (toponym, place) pair; and

(2) if a toponym is identified within the document that has an associated place that is geographically related to the place of the selected (toponym, place) pair, boosting the value of the confidence for the selected (toponym,place) pair.

11. The computer-implemented method of claim 10, wherein determining if another toponym is present within the document that has an associated place that is geographically related to the place of that (toponym, place) pair involves identifying another toponym that has an associated geographic region that encompasses the place of the selected (toponym, place) pair.

12. The computer-implemented method of claim 10, wherein determining if another toponym is present within the document that has an associated place that is geographically related to the place of that (toponym, place) pair involves identifying another toponym that has an associated place that is geographically nearby the place of the selected (toponym, place) pair.

1   13. The computer-implemented method of claim 12, further comprising computing a

2 geographical distance between the place associated with the identified toponym and the place

3 of the selected (toponym,place) pair.

1   14. The computer-implemented method of claim 13 wherein boosting involves

2 calculating an adjustment value by computing an adjustment boosting function with the

3 computed geographical distance as an input variable, said adjustment function being

4 monotonically decreasing for increasing values of the input variable.

1   15. The computer-implemented method of claim 14 wherein boosting involves

2 deriving an initial boosting value from input including the calculated adjustment value.

1   16. The computer-implemented method of claim 14 wherein boosting involves

2 applying a sigmoid function to the derived initial boosting value to compute a final boosting

3 value and modifying the value of the confidence for the selected (toponym,place) pair by an

4 amount determined by the final boosting value.

1   17. The computer-implemented method of claim 11 further comprising:

2   performing steps (1) and (2) for each (toponym,place) pair among the plurality of

3 (toponym,place) pairs to generate modified values for the confidences for the plurality of

4 (toponym,place) pairs; and

5   using the modified values to rank documents according to their relevance to a search

6 query.

1   18. A method of evaluating relevance of a plurality of documents to a search query

2 that includes both text and geographic place terms, said method comprising:

3   for a selected document among the plurality of documents,

4   (1) computing a textual term relevance score corresponding to the text terms in the

5 query;

6   (2) computing a geo-relevance score corresponding to the geographic terms in the

7 query; and

8   (3) combining the computed textual term relevance score and the computed geo-

9 relevance score to derive an overall relevance score for that document,

24

10   wherein computing the geo-relevance for the selected document involves identifying

11   a plurality of (toponym,place) pairs that is associated with the selected document, and for

12   each identified (toponym,place) pair, obtaining and using a value for a confidence that the

13   toponym of the (toponym,place) pair refers to the place.